

## REMARKS

No claims have been added, cancelled, or amended. Hence, claims 1 – 34 are pending in the application.

## SUMMARY OF REJECTIONS/OBJECTIONS

The specification has been objected to because it contains hyperlinks. Applicant has amended the application to resolve this objection. Removal of the objection is respectfully requested. Applicant has submitted FIG. 6, a new drawing. It contains material previously embedded within the specification. The Examiner's approval of the FIG. 6 is respectfully requested.

Claims 1 – 14, 17 – 19, and 26 – 34 are rejected under 35 USC 103(a) as being unpatentable over U.S. Patent No. 5,835,905, herein *Pirolli*, in view of U.S. Patent No. 5,906,422, herein *Prasad*.

Claims 15 – 16 are rejected under 35 USC 103(a) as being unpatentable over *Pirolli* and *Prasad*, and in further view of U.S. Patent No. 6,282,549, herein *Hoffert*.

Claim 20 is rejected under 35 USC 103(a) as being unpatentable over *Pirolli* and *Prasad* in view of U.S. Patent No. 6,128,606, herein *Bengio*.

Claims 21 – 25 are rejected under 35 USC 103(a) as being unpatentable over *Pirolli* and *Prasad* in view of U.S. Patent No. 6,389,436, herein *Chakrabarti*.

## REJECTION OF CLAIMS 1 AND 34 UNDER USC 102

Claims 1 and 34 recite:

**determining how strongly each document of said plurality of documents  
corresponds to each of said plurality of categories by determining**

similarity between said each document and the training documents that belong to the training set of said category....

Claims 1 and 34 recite limitations not disclosed or suggested by the cited art. Among these limitations are “**determining how strongly each document of said plurality of documents corresponds to each of said plurality of categories** by determining similarity between said each document and the documents that belong to the training set of said category.”

The Office Action has based the rejections of claims 1 and 34 on *Pirolli*. Applicant admits that *Pirolli* teaches (1) to categorize a set of documents, in the form of pages, according to “classification characteristics”, and (2) to determine textual similarity between documents to categorize a document. However, Applicant is not attempting to claim only these features. Rather, Applicant is claiming to use the similarity between a document and a particular set of documents (i.e., training set), which have been established as belonging to a category, to determine the correspondence between the document and the category.

*Pirolli* teaches that documents are categorized into functional categories which are “designed by someone (application designer, webmaster, end user), in contrast to being automatically induced.” (col. 8, lines 34 - 36). A number of characteristics are used to classify documents. Only one of these characteristics are based on similarity between a document and a particular set of documents. That characteristic is csim; “csim, [is] the textual similarity of the item to it's children based upon previous SCA calculation (column 508).”

*Pirolli* further teaches that text similarity is used to determine whether a page belongs to the category of head page (e.g., home page) (col. 9, lines 14 – 24).

For Head Nodes (classification criteria 601), being the first pages of a collection of documents with like content, it is expected that such pages will have high text similarity between itself and its children, and would have a high average depth of its children, and that it would be more likely to be an entry point based upon actual user navigation patterns.

Thus, at best, *Pirolli* teaches that text similarity between a page and the children of the page is used to determine the correspondence between the page and the category of home page. However, this is not a category to which the set of children have been established as belonging. The claims, on the other hand, require the feature of using similarity between a document and a particular set of documents established as belonging to a category to determine the correspondence between the document and the category.

In fact, *Pirolli* seems to teach against such a feature because of the types of functional categories it discloses. For example, head node is a category which includes documents in which text similarity between the documents in this category is of little relevance. Examples of a set of documents that could be established in this category are Yahoo's home page, Google's home page, and the USPTO home page. It would seem that text similarity between these pages and another page would have very little relevance to whether the other page is a home page.

*Prasad* also fails to teach the claimed feature of using similarity between a document and another set of documents established as belonging to a category to determine the correspondence between the document and the category. Presumably, the Office Action has equated a document as claimed to a document at a data source and a training set as claimed to a sample of documents from a data source. Even if the training set taught by *Prasad* can be equated to the training set claimed, *Prasad* nevertheless fails to teach the claimed feature.

*Prasad* teaches that rule induction is applied to the training set to generate rules that are used to determine what source to direct queries. (col. 3, line 66 – col. 4, line 16). While *Prasad* teaches that training sets are used as input for rule induction, no teaching in *Prasad* suggests training sets are used determine the correspondence between a document and the category to which the training set belongs by determining the similarity between the document and the training set.

#### PENDING CLAIMS

The pending claims not discussed so far are dependant claims that depend on an independent claim that is discussed above. Because each of the dependant claims include the limitations of claims upon which they depend, the dependant claims are patentable for at least those reasons the claims upon which the dependant claims depend are patentable. Removal of the rejections with respect to the dependant claims and allowance of the dependant claims is respectfully requested. In addition, the dependent claims introduce additional limitations that independently render them patentable. Due to the fundamental difference already identified, a separate discussion of those limitations is not included at this time.

The Examiner is respectfully requested to contact the undersigned by telephone if it is believed that such contact would further the examination of the present application.

For the reasons set forth above, Applicant respectfully submits that all pending claims are patentable over the art of record, including the art cited but not applied.

Accordingly, allowance of all claims is hereby respectfully solicited.

Respectfully submitted,

HICKMAN PALERMO TRUONG & BECKER LLP



Marcel K. Bingham  
Reg. No. 42,327

Dated: April 23, 2003

1600 Willow Street  
San Jose, CA 95125  
Telephone No.: (408) 414-1080 ext.206  
Facsimile No.: (408) 414-1076

**CERTIFICATE OF MAILING**

I hereby certify that this correspondence is being deposited with the United States Postal Service as first class mail in an envelope addressed to: Box Non-Fee Amendment, Commissioner for Patents, Washington, D.C. 20231

on April 23, 2003

by 